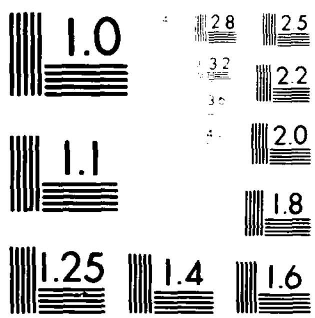
AD-A097 113 AIR FORCE ENVIRONMENTAL TECHNICAL APPLICATIONS CENTER--ETC F/6 4/1
SIZING AN ENVIRONMENTAL DATA SET,(U)
APR 78 E C TAYLOR
UNCLASSIFIED USAFETAC-PR-78-004 SBIE-AD-E850 035 NL

1 OF 1

END
DATE FILMED
5-81
DTIC



MICROCOPY RESOLUTION TEST CHART

ADF 850 035

UNITED STATES AIR FORCE
AIR WEATHER SERVICE (MAC)

USAF ENVIRONMENTAL
TECHNICAL APPLICATIONS CENTER

SCOTT AIR FORCE BASE, ILLINOIS 62225

1 LEVEL III

AD A097113

USAFETAC-PR-78-004

SIZING AN ENVIRONMENTAL DATA SET

by

Elden C. Taylor, Major, USAF

April 1978

Approved for public release; distribution unlimited.

REVIEW AND APPROVAL STATEMENT

This report is approved for public release. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS).

This technical report has been reviewed and is approved for publication.

Elden C. Taylor

ELDEN C. TAYLOR
Maj, USAF

John F. Morrissette

JOHN F. MORRISSETTE
LtCol, USAF
Reviewing Officer

FOR THE COMMANDER

Walter S. Burgmann

WALTER S. BURGMANN
Scientific and Technical
Information Officer (STINFO)

UNCLASSIFIED 04/01

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
USAFETAC-PR-78-004

2. GOVT ACCESSION NO.
AD-A097113

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Sizing an Environmental Data Set

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Elden C. Taylor, Maj, USAF

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
USAF Environmental Technical Applications Center
Scott AFB, IL 62225

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
USAFETAC Project 1219

11. CONTROLLING OFFICE NAME AND ADDRESS
USAF Environmental Technical Applications Center
Scott AFB, IL 62225

12. REPORT DATE
April 1978

13. NUMBER OF PAGES
4

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Climatology
Data set
Environmental data set
Data set size
Worldwide Military Command and Control System (WWMCCS)

*Mathemat[illegible]
*Mathem[illegible]
Mathem[illegible]
Data bases
*Data management

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents a basic equation and three derivative equations that can be used to size environmental data sets. These equations were formulated to document mass storage requirements for USAFETAC's environmental support to the Worldwide Military Command and Control System (WWMCCS). With minor modifications, these equations have a much wider application. They are applied to a specific example taken from a request for environmental support.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SIZING AN ENVIRONMENTAL DATA SET

Introduction

This report presents a basic equation and three derivative equations that can be used to size environmental data sets. These equations were formulated to document mass storage requirements for USAFETAC's environmental support to the Worldwide Military Command and Control System (WWMCCS). With minor modifications, these equations have a much wider application. They are applied to a specific example taken from a request for environmental support.

A set of diagnostic equations which can be used in quantitative quality control applications are also included.

The Basic Equation

Each climatological study begins with weather observations taken at specific times and locations and containing a measure of meteorological elements. These observations are collected at numerous locations and times for varying periods of record (POR) and stored for later summarizing, analysis, or both.

Equation (1) provides an estimate of the amount of storage that will be required for a particular data set.

$$S_o = (P)(L)(E)(Y) \tag{1}$$

where S_o = storage required for the record of observations

P = number of meteorological elements

L = number of observation sites

Y = period of record (POR) parameter

E = storage parameter

The POR parameter, Y, can be expanded to

$$Y = (y)(h)(d)$$

where y = number of years of record

h = number of observations per day

d = number of days per calendar unit

d = 365 (days/year) for most applications, but

d = 31 (days/January) for some applications.

The storage parameter, E, can be expressed in terms dictated by the storage medium, e.g., observations per page for paper storage or observations per microfiche. This paper will use

$$E = 8B$$

where B = bytes per parameter, and

8 = bits per byte (IBM 360/44).

Making the above substitutions in Equation (1)

$$S_o = (P)(L)(8)(B)(y)(h)(365) \tag{1a}$$

which is the form that is useful at the USAF Environmental Technical Applications Center (USAFETAC).

Derived Equations

Climatology requires summarization. Equation (2) applies to summarized data.

$$S_s = (\overline{P})(\overline{L})(\overline{h})(A)(E) \tag{2}$$

where S_s = storage required for the record of summarized observations

$\overline{P}$ = summarized parameter of measurement of the meteorological element

$\overline{L}$ = number of observation sites summarized

$\overline{h}$ = number of times (e.g., hours) for which $\overline{P}$ is computed

A = number of periods (e.g., months) for which $\overline{P}$ is computed

E = as in Equation (1).

We can carry the derivation one step further. If we need the elements summarized in combination, e.g., ceiling/visibility, wind chill factors (wind/temperature), or by location (when is the temperature the same at two or more locations), or in combination of elements at two or more locations (probability of favorable takeoff conditions at location A <u>and</u> favorable landing conditions at location B 6 hours later - a mission success indicator), we can use Equations (3) and (4) to estimate the storage required.

Equation (3) allows one to estimate the storage required for "any combination of elements." Equation (4) is used for a combination of elements at multiple locations.

$$S_C = (C)(\overline{L})(\overline{h})(A)(E) \tag{3}$$

$$S_{CL} = (C)(L')(\overline{h})(A)(E) \tag{4}$$

where

$$C = \frac{\overline{P}!}{(\overline{P}-N)!\,N!} \qquad L' = \frac{\overline{L}!}{(\overline{L}-M)!\,M!}$$

S_C = storage required for the record of combinations of summarized observations

S_{CL} = storage required for the record of multiple-location combinations of summarized observations

C = number of meteorological elements taken N at a time

L' = number of locations taken M at a time

When $N = 1$, $C = \overline{P}$ and when $M = 1$, $L' = \overline{L}$.

Diagnostic Equations

To test "completeness" of a summarized data set we can use Equations (5) and (6).

$$D_E = (C)(L)(Y) \tag{5}$$

$$D_O = (C)(L)(Y) \qquad (6)$$

where D_E = number of observations expected in a complete data set

D_O = number of observations found in the data set

C = number of combinations of elements

Y = (y)(h)(d)

A ratio of D_O and D_E gives a measure of completeness and

$$(100)\ \frac{D_O}{D_E} = K \qquad (7)$$

where K = the percentage "complete."

For example, in using the Revised Uniform Summary of Surface Weather Observations (RUSSWO) for Altus AFB, OK we find that 15,463 observations (D_O) were used to compute January all hours ceiling versus visibility for the POR 1944-45 and 1954-72. Using Equation (5) we find C = 1, L = 1, y = 21, h = 24, and d = 31 (days/January). Then

$$D_E = (21)(24)(31) = 15{,}624$$

and

$$(100)\ \frac{D_O}{D_E} = (100)\left(\frac{15463}{15624}\right) = 98.97\% \text{ complete}$$

<u>Practical Example</u>

USAFETAC received a proposal for support to the US Army requesting "all parameters used in Army operations for up to 500 locations. These parameters may be needed in any combination." A list of 23 elements followed. Using Equation (1a) with P = 23, L = 500, h = 24 (hours/day), y = 10 (years), and B = 1, we find $S_o = (8.0592)(10^9)$ bits. This is about thirty 2400-foot magnetic tapes written at a density of 1600 bits per inch.

To satisfy this request USAFETAC would need to summarize by month, plus an annual value, all of the parameters for all locations and for all hours. Analytically, then, the assumptions are $P = \bar{P}$ (23), $L = \bar{L}$ (500), $h = \bar{h}$ (24), y = 10, E = E, A = 13 (monthly values plus annual). From Equation (2) we find:

$$S_s = (23)(500)(24)(13)(8) = (2.8704)(10^7)$$

Or, taking a ratio of S_o/S_s:

$$\frac{S_o}{S_s} = \frac{(P)(L)(h)(10)(365)(E)}{(\bar{P})(\bar{L})(\bar{h})(13)(E)}$$

If we allow 365/13 to be approximated by 30, we find that

$$S_o = 300\ S_s \qquad (8)$$

Equation (8) says that 300 tapes of observations can be reduced to one tape of summarized data under the stated assumptions. This ratio approximates the current procedures at USAFETAC.

Moving to Equation (3), we find that the total number of combinations of P parameters taken 1, or 2, ..., or N at a time can be expressed by

$$\sum_{i=1}^{N} C_i = 2^N - 1$$

When N = 23, C_{23} = 8,388,607. Table 1 shows a step-by-step increase in the required storage as more and more combinations are considered. The TOTALS line in Table 1 show that 300 tapes of "raw" observations of 23 elements can generate 364,855 tapes of combinations of elements, and there remains 300 tapes of observations for a total storage requirement of 365,155 tapes.

If we take stations three at a time from a population of 500 stations, we find L' = 20,708,500. Apply this to the total in Table 1 and, using Equation (4), we find that we can easily generate $(6)(10^{12})$ tapes of summaries for any combination of elements for any combination of three stations from a population of 500 stations.

Table 1. Computed Storage Requirements.

N	C	RATIO $\frac{S_e}{S_C}$	TAPES S_C PER 300 TAPES S_o
N = 1 or 22	23	300.	1
N = 2 or 21	253	27.22273	11
N = 3 or 20	1,771	3.89610	77
N = 4 or 19	8,855	0.78022	385
N = 5 or 18	33,649	0.20506	1,463
N = 6 or 17	100,947	0.06835	4,390
N = 7 or 16	245,157	0.02815	10,658
N = 8 or 15	490,314	0.01407	21,322
N = 9 or 14	817,190	0.00844	35,545
N = 10 or 13	1,144,066	0.00603	49,751
N = 11 or 12	1,352,078	0.00510	58,824
N = 23	1	6,900.	1
TOTALS	$\sum_{i=1}^{23} C_i = 8,388,607$		364,855

Note

USAFETAC has, at Scott AFB, IL and Asheville, NC, about 25,000 magnetic tapes dedicated to storing observations. These data include about 9400 stations reporting surface data and 2000 stations reporting upper-air data for which "suitable" periods of records exist. An additional 20,000 magnetic tapes contain summaries and other derived information.

DATE
ILMED
-8